

DOCUMENT RESUME

ED 217 695

FL 012 935

AUTHOR Canale, Michael
 TITLE The Method Effect in Communicative Testing.
 INSTITUTION Public Service Commission of Canada, Ottawa (Ontario).
 PUB DATE 81
 NOTE 15p.
 JOURNAL CIT Medium; v6 n4 p43-55 Dec 1981
 EDRS PRICE MF01/PC01 Plus Postage.
 DESCRIPTORS *Communicative Competence (Languages); *Language Tests; Methods Research; Second Language Instruction; Speech Communication; Test Construction; *Test Validity

ABSTRACT

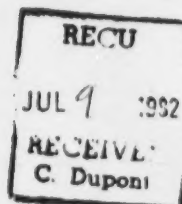
A focus on test validity includes a consideration of the way a test measures that which it proposes to test; in other words, the validity of a test depends on method as well as content. This paper examines three areas of concern: (1) some features of communication that test method should reflect, (2) the main components of method, and (3) some examples of how method may affect a testee's performance. Seven features of communication are proposed as a checklist with which test methods may be assessed for their degree of correspondence to real communication. To illustrate how influential method can be in testing communication, five method factors are described. These are mode, response type, administration procedures, environment, and scoring procedures. A sampling of evidence is then given to illustrate how each method factor can affect test scores. Finally, suggestions are given for addressing the problem of method effect in communicative test development. One way is to try to develop a very practical indirect communicative test that would be a good predictor of performances on a direct and comprehensive test. A direct test would be one that seeks to make test methods correspond as directly as possible to the demands likely to be made on the testee in authentic communication situations.
 (AMH)

 * Reproductions supplied by EDRS are the best that can be made *
 * from the original document. *



U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

REPRODUCTION RELEASE (Specific Document)



I. DOCUMENT IDENTIFICATION

Title: The Method Effect in Communicative Testing Date: July 9, 1982
Author(s): Michael Canale
Corporate Source (If appropriate): _____

II. REPRODUCTION RELEASE

In order to disseminate as widely as possible timely and significant materials of interest to the educational community, documents announced in the monthly abstract journal of the ERIC system, Resources in Education (RIE), are usually made available to users in microfiche and paper copy (or microfiche only) and sold through the ERIC Document Reproduction Service (EDRS). Credit is given to the source of each document, and, if reproduction release is granted, one of the following notices is affixed to the document.

If permission is granted to reproduce the identified document, please CHECK ONE of the options and sign the release below.

CHECK
HERE



Microfiche
(4" x 6" film)
and paper copy
(8 1/2" x 11")
reproduction

"PERMISSION TO REPRODUCE THIS
MATERIAL HAS BEEN GRANTED BY

(PERSONAL NAME OR ORGANIZATION)
Michael Canale
AS APPROPRIATE]

MICHAEL CANALE
TO THE EDUCATIONAL RESOURCES
INFORMATION CENTER (ERIC)."

OR



Microfiche
(4" x 6" film)
reproduction
only

"PERMISSION TO REPRODUCE THIS
MATERIAL IN MICROFICHE ONLY
HAS BEEN GRANTED BY

(PERSONAL NAME OR ORGANIZATION)

AS APPROPRIATE]

TO THE EDUCATIONAL RESOURCES
INFORMATION CENTER (ERIC)."

Documents will be processed as indicated provided reproduction quality permits. If permission to reproduce is granted, but neither box is checked, documents will be processed in both microfiche and paper copy.

SIGN
HERE

"I hereby grant to the Educational Resources Information Center (ERIC) nonexclusive permission to reproduce this document as indicated above. Reproduction from the ERIC microfiche by persons other than ERIC employees and its system contractors requires permission from the copyright holder. Exception is made for non-profit reproduction of microfiche by libraries and other service agencies to satisfy information needs of educators in response to discrete inquiries."

Signature: Michael Canale Printed Name: MICHAEL CANALE
Organization: O.I.S.E. Position: ASSISTANT PROFESSOR
Address: 252 BLOOR ST. W. Tel. No.: (416) 923-6641
TORONTO, ONT. CANADA Zip Code: M5S 1V6

III. DOCUMENT AVAILABILITY INFORMATION (Non-ERIC Source)

If permission to reproduce is not granted to ERIC, or, if you wish ERIC to cite the availability of the document from another source, please provide the following information regarding the availability of the document. (ERIC will not announce a document unless it is publicly available, and a dependable source can be specified. Contributors should also be aware that ERIC selection criteria are significantly more stringent for documents which cannot be made available through EDRS.)

Publisher/Distributor: _____
Address: _____
Price Per Copy: _____ Quantity Price: _____

IV. REFERRAL TO COPYRIGHT/REPRODUCTION RIGHTS HOLDER

If the right to grant reproduction release is held by someone other than the addressee, please provide the appropriate name and address:

THE METHOD EFFECT IN COMMUNICATIVE TESTING

Michael Canale is an Assistant Professor at
the Ontario Institute for Studies in Education.
He can be reached in Toronto at (416) 923-6641.

0. Introduction

The general focus of this paper is on the validity of tests of communication. By validity is meant how well a test measures what it should measure; for example, a valid communicative test is one that adequately measures the knowledge and skills crucial to the communicative use of language. As understood here, then, the validity of a test depends on what the test tries to measure and how it tries to measure this, or on what are often referred to as test content and test method respectively.

In the area of communicative testing, these concerns with content and method have been addressed in some very exciting research (e.g. Bachman and Palmer 1981a, b; and Oller 1981). But there are reasons to believe that the importance of method has nonetheless been underestimated in much recent work, including my own. One reason is that in many current language tests and handbooks on communicative testing, the components of method and their potential effect on performance do not receive much explicit attention. Another is the realization that, especially in communicative testing, a test is in fact not so much an objective tool for collecting data as it is a complex event for creating data (cf. Troike 1981; Walkerdine and Sinha 1981). It is for this last reason that many disciplines (e.g. primatology--cf. Rosen 1974) recognize fundamental distinctions among data according to methods of observation.

In view of these reasons, the specific focus of this paper is on the importance of various components of measurement method for communicative testing. The following three sections examine briefly (1) some features of communication that test method should reflect, (2) the main components of method, and (3) some examples of how method may affect a testee's performance. A con-

"PERMISSION TO REPRODUCE THIS
MATERIAL HAS BEEN GRANTED BY

Michael Canale

TO THE EDUCATIONAL RESOURCES
INFORMATION CENTER (ERIC)."

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)

* This document has been reproduced as
received from the person or organization
originating it.

☐ Minor changes have been made to improve
reproduction quality.

• Points of view or opinions stated in this docu-
ment do not necessarily represent official NIE
position or policy.

ED217695

FL012935

cluding section suggests some lines of further research. It should be noted that my concern with method over content here is for purposes of presentation, and is in no way meant to downplay the important and complex issues that remain in determining the content of a valid communicative test.

1. Features of communication

The most useful description of communication that I have examined is that of Morrow (1977). The seven features of communication he proposes can serve as a checklist with which test methods may be assessed for their degree of correspondence to real communication. If a test method shows few of these features, then one may doubt the validity of the test for measuring communicative language use. These features can be formulated as follows. Communication:

- (a) is interaction-based in that communication skills are normally both acquired and used in social interaction;
- (b) involves unpredictability and creativity in form and message;
- (c) takes place in discourse and sociocultural contexts which provide constraints on appropriate language use and also clues as to correct interpretations of utterances;
- (d) is carried out under limiting psychological and other conditions such as memory constraints, fatigue, and interfering noise;
- (e) always has a purpose (for example, to establish social relations, to encourage, or to complain) and usually has more than one purpose;
- (f) involves authentic and personal language as opposed to textbook-contrived utterances; and
- (g) is judged as successful or not on the basis of actual outcomes. (For example, communication could be judged successful in the case of a nonnative English speaker who was trying to find the train station in Toronto, uttered the ungrammatical sentence 'How to go train?' to a passer-by, and was given directions to the train station.)

Note that this description of communication applies equally well to reading, writing, listening, and speaking. For further discussion and references regarding the nature of communication, see Canale (1981a).

2. The main components of test method

A communicative test may be developed and administered for a number of different purposes such as placement in a language program, monitoring student achievement, diagnosis of difficulties, and hiring and classification of employees. Regardless of the purpose, one is normally interested in obtaining accurate information about a testee's proficiency in communicating, or communicative competence. In testing terminology, one could refer to this communicative competence as a trait (or group of traits). A test score is intended to give accurate information about this trait (for example, how one person's communicative competence compares with another's).

In a very instructive article, Bachman and Palmer (1981c) point out that a test score reflects not only the trait to be measured but also the method of measurement, error in measurement, and the uniqueness or specificity of measurement (idiosyncratic properties of any test). The goal of test validation is not only to determine how much of a score is due to each but also to minimize the effects of method, error, and uniqueness while maximizing the effect of trait. The discussion that follows is concerned only with method effects.

To illustrate just how influential method can be in testing communication, the overview of language testing presented in Figure 1 may be useful. This overview is based on the work of Bachman and Palmer (1981c) and is not intended to replace the presentation there. The five method factors identified in Figure 1 can be described as follows.

(a) Mode. This refers to the language skill areas of reading, writing, listening, and speaking which may be isolated or combined in a test. Also included are the mode in which the test is presented to the testee and the mode in which the response is required.

(b) Response type. The main distinction here is between formats that require a more discrete-point response (e.g., multiple choice items and fill-in-the-blanks) versus ones requiring more integrative response (e.g., telling a story or engaging in a conversation). The distinction between direct and indirect testing may also be thought of here (cf. Clark 1979 and Wesche 1981 for discussion).

(c) Administration procedures. The concern here is with how well instructions are understood (e.g., the language in which they are presented, their complexity, and their clarity) and with the time allowed the testee to complete the test.

(d) Environment. This includes setting (e.g., comfort and how much setting corresponds to an authentic context in which to perform test tasks) and personnel (number, sex, familiarity to the student, and degree to which an interviewer encourages communication, for example).

(e) Scoring procedures. These may be objective or subjective; if subjective, the concern is with the judges (e.g., self-evaluation, standards, and consistency) and scoring criteria (e.g., selection, clarity, and weighting relative to one another).

3. Examples of the method effect

There is growing evidence that each method factor distinguished in Figure 1 can affect test scores in little understood ways. The following is just a sampling of such evidence.

Mode. Singer (1977) and Troike (1981) discuss the strong relationship between performance on reading comprehension tasks and intelligence test tasks; Oller and Hinofotis (1980) and Cummins (1981) discuss findings that suggest less of a relationship between performance on oral interaction tasks and performance on intelligence test tasks and academic achievement tasks. Bachman and Palmer (1981a, b) report on the differences in measurement associated with different methods of evaluating communicative competence in speaking and in reading.

Response type. Shohamy (1980) has observed an interaction between attitudes toward different tests (namely cloze and oral interview) and performance on those tests. Alderson (1980) discusses the lack of consistent performance that emerges among native and non-native speakers when different forms of cloze tests are administered. Jones (1981) and Oller (1979) argue strongly for use of integrative type tasks when assessing oral proficiency and language proficiency, respectively.

Administration. The language in which test instructions are given has been an important issue in work on both the Ontario Assessment Instrument Pool for French as a Second Language (cf. McLean 1980) and English as a Second Language (or Anglais--J.-M. Joly, personal communication). Carroll (1961) has argued that the more time a testee is given to focus on individual, discrete aspects of language, the less such tasks correspond to authentic language-use demands.

Environment. Upshur and Palmer (1974) report findings suggesting that performance on language proficiency tests in a classroom setting is not a good predictor of performance on communicative tasks outside the classroom. As concerns personnel, Troike (1981) reports a 30% increase in the performance of black students when administered a test by computer rather than by a white adult; only marginal differences were found between the scores of white students in this study. A similar phenomenon is discussed by Labov (1969). See also Lowe (1981) for discussion of the importance and complexity of training for interviewers for the Foreign Service Institute Oral Proficiency Interview.

Scoring. Clark (1975) reports reliability coefficients of as low as .31 among scorers of the MLA-Cooperative Speaking Tests for French. Clifford (1981) observes possible differences in standards in scoring according to the language being tested in the Foreign Service Institute Oral Proficiency Interview. Oskarsson (1981) discusses the low reliability of judgements in self-evaluation of language proficiency. Concerning scoring criteria, Cummins et al. (1981), Hinofotis (1981) and Lepicq (1980) have sought to identify the relative importance of a variety of criteria in evaluations of oral communication skills. Evans (1981) has discussed the difficulty of keeping various criteria

apart in the evaluation of writing, and has also explored the possible 'halo effect' of handwriting on scorers' judgements of overall quality of student writing.

4. Conclusion

Sketchy and tentative as they may be, the above examples nonetheless serve to suggest some of the ways in which each component of test method may affect performance on a communicative test. Assuming that there is a method effect reflected in test scores and that it is relatively little understood, one cannot ignore it and still claim that a given communicative test is valid. How then might one address it in communicative test development?

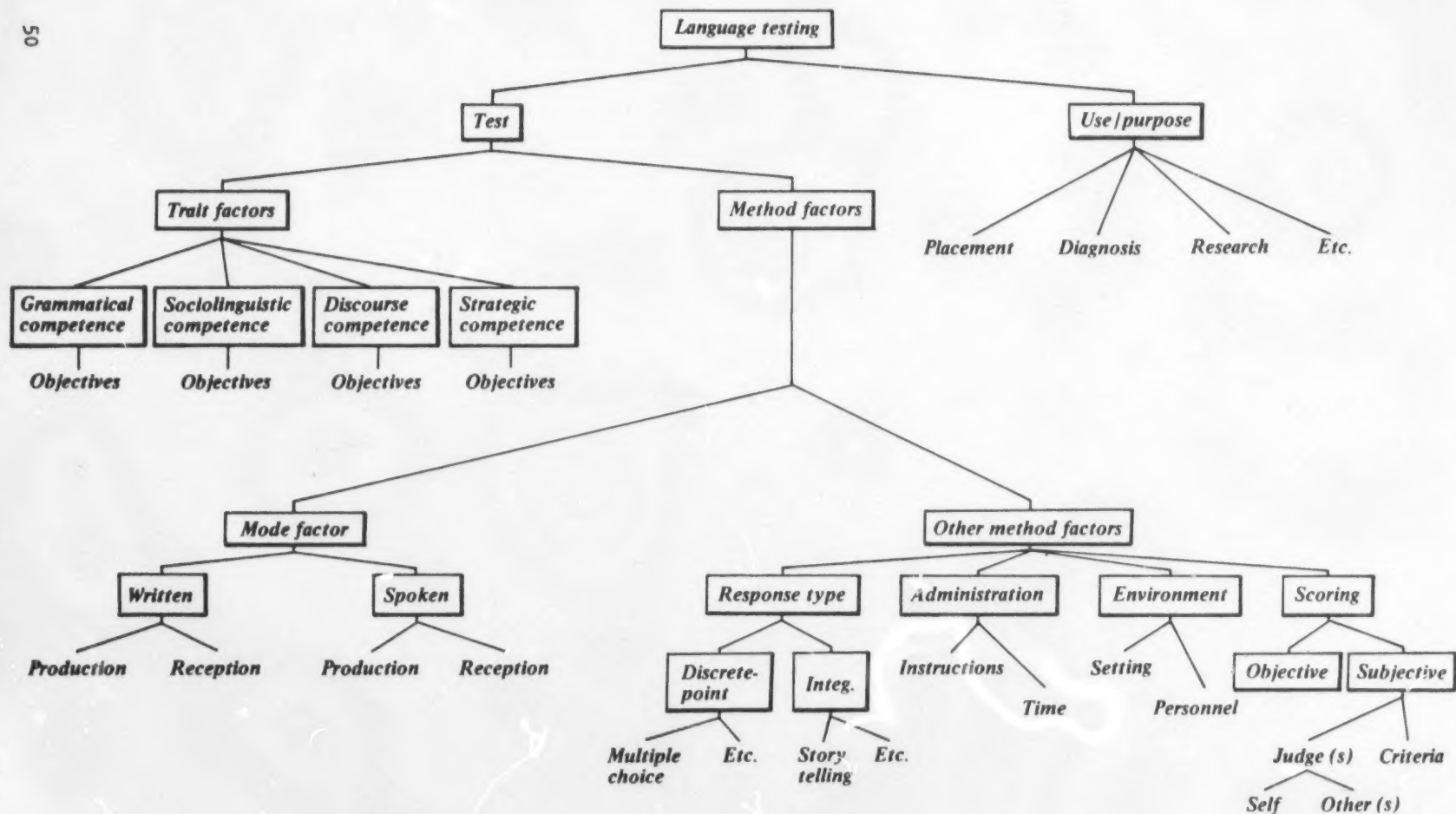
One straightforward way is to seek to make test methods correspond as directly as possible to the demands made or likely to be made on the testee in authentic communication situations. For example, Hinofotis (1981) reports on the evaluation of the oral language skills of nonnative teaching assistants at U.C.L.A. in which method factors such as test tasks, environment, judges, and scoring procedures were selected so as to reflect directly the actual demands made on the teaching assistants. Such an approach cannot, of course, eliminate method effects from communicative testing any more than one can eliminate such factors as IQ, world knowledge, personal values, and personality from authentic communication. But a direct testing approach can help to assure that the components of the testing process and their effects match those of authentic communication.

However, by responding to the problem of method effect in the manner just suggested, the language tester would run the risk of developing a test that is not always practical and hence not always usable. That is, the development, administration, and scoring of a direct test of communication may pose insurmountable financial and administrative obstacles to decision-makers. To overcome this difficulty, a second approach may be preferred. Thus one might try to develop a very practical indirect communicative test that is a good predictor of performances on a direct and comprehensive (less practical) test. Test validation and development frequently proceed in this way; for example, selection

of possible responses (distractors) in a multiple-choice test might be based on testees' actual responses to an open-end form of the test (see Morrow 1977 and Wesche 1981 for further discussion). Additional suggestions for making communicative tests more practical can be found in Canale (1981b), for example. A serious problem in the use of indirect communicative tests is raised by Clark (1972) and Jones (1981): such indirect assessment procedures as pencil-and-paper tests and tape-recorded speaking and listening tests cannot have the same psychological and instructional impact on testees as do more direct tests requiring more authentic communicative interaction.

Whatever route is chosen to deal with the method effect, the warning sounded by Clark and Jones must not be ignored. Concern with the effect of method on a test score, though important, must not override concern with the effect of any test and testing program on the individuals under study.

Figure 1. A schematic outline of concerns in language testing.



REFERENCES

- ALDERSON, J.C. 1980. Native and Nonnative Speaker Performance on Cloze Tests. Language Learning 30.1.
- BACHMAN, L. and A.S. Palmer. 1981a Basic Concerns in Test Validation. In J.A.S. Read, (Ed.) Directions in Language Testing. Singapore: RELC.
- BACHMAN, L. and A.S. Palmer. 1981b. The Construct Validation of the FSI Oral Interview. Language Learning 31.
- BACHMAN, L. and A.S. Palmer. 1981c. Some Comments on the Terminology of Language Testing. Paper presented at the Language Proficiency Assessment Symposium, Airlie House, Virginia, March. To appear in the proceedings.
- CANALE, M. 1981a. From Communicative Competence to Communicative Language Pedagogy. In J.C. Richards and R. Schmidt, (Eds.). Language and Communication. London: Longman.
- CANALE, M. 1981b. Communication: How to Evaluate it? Keynote address presented at the Annual Meeting of the Canadian Association of Applied Linguistics, Carleton University, Ottawa, May. To appear in the proceedings in the Bulletin of the Canadian Association of Applied Linguistics.
- CARROLL, J.B. 1961. Fundamental Considerations in Testing for English Language Proficiency of Foreign Students. In Testing the English Proficiency of Foreign Students. Washington, D.C.: Center for Applied Linguistics. Reprinted in H.B. Allen and R.N. Campbell, (Eds.). 1972. Teaching English as a Second Language: A Book of Readings. New York: McGraw-Hill.

- CLARK, J.L.D. 1972. Foreign Language Testing: Theory and Practice. Philadelphia: The Center for Curriculum Development.
- CLARK, J.L.D. 1975. Theoretical and Technical Considerations in Oral Proficiency Testing. In R.L. Jones and B. Spolsky, (Eds.). Testing Language Proficiency. Arlington, Va.: Center for Applied Linguistics.
- CLARK, J.L.D. 1979. Psychometric Considerations in Language Testing. In B. Spolsky, (Ed.). Advances in Language Testing series, 2: Approaches to Language Testing. Washington, D.C.: Center for Applied Linguistics.
- CLIFFORD, R. 1981. Oral Proficiency Performance Profiles and Global Rating. Paper presented at the Pre-Conference on Oral Proficiency Assessment, Georgetown University Round Table, Washington, D.C., March. To appear in the proceedings.
- CUMMINS, J. 1981. Is Academic Achievement Distinguishable from Language Proficiency? In J.W. Oller, Jr., (Ed.). Issues in Language Testing Research. Rowley, Mass.: Newbury House.
- CUMMINS, J. et al. 1981. Linguistic Interdependence among Japanese Immigrant Students. Paper presented at the InterAmerica Language Proficiency Assessment Symposium, Airlie House, Virginia, March. To appear in the proceedings.
- EVANS, P.J.A. 1981. The Analysis of Writing Assessment: the O.I.S.E. Writing Evaluation Project in Ontario 1977-1980. Paper presented at the Annual Conference of the Canadian Council of Teachers of English, Vancouver, May. Mimeo.

- HINOFOTIS, F.E. 1981. The Structure of Oral Communication in an Educational Environment: A Comparison of Factor Analytic Rotational Procedures. In J.W. Oller, Jr., (Ed.). Issues in Language Testing Research. Rowley, Mass.: Newbury House.
- JONES, R.L. 1981. Scoring Procedures in Oral Language Proficiency Testing. In J. Read, (Ed.). Directions in Language Testing. Singapore: RELC.
- LABOV, W. 1969. The Logic of Nonstandard English. Georgetown Monographs on Language and Linguistics 22.
- LEPICQ, D. 1980. Aspects théoriques et empiriques de l'acceptabilité linguistique: Le cas du français des élèves des classes d'immersion. Unpublished doctoral thesis, University of Toronto.
- LOWE, P. 1981. Components of Oral Examiner/Interviewer Training. Paper presented at the Pre-Conference on Oral Proficiency Assessment, Georgetown University Round Table, Washington, D.C., March. To appear in the proceedings.
- McLEAN, L.D. 1980. Separating the Effect of More Hours of Instruction from General Language Development on Achievement in French as a Second Language. Paper presented at the Fourth International Symposium on Educational Testing, Antwerp, Belgium, June. Mimeo.
- MORROW, K.E. 1977. Techniques of Evaluation for a Notional Syllabus. London: The Royal Society of Arts.
- OLLER, J.W., Jr. 1979. Language Tests at School. A Pragmatic Approach. London: Longman, Inc.

OLLER, J.W., Jr. (Ed.). 1981. Issues in Language Testing Research. Rowley, Mass.: Newbury House.

OLLER, J.W., Jr. and F.B. Hinofotis. 1980. Two Mutually Exclusive Hypotheses about Second Language Ability: Indivisible and Partially Divisible Competence. In J.W. Oller, Jr. and K. Perkins, (Eds.). Research in Language Testing. Rowley, Mass.: Newbury House.

OSKARSSON, M. 1981. Subjective and Objective Assessment of Foreign Language Performance. In J.A.S. Read, (Ed.). Directions in Language Testing. Singapore: RELC.

READ, J.A.S., ed. 1981. Directions in Language Testing. Singapore: RELC.

ROSEN, S. 1974. Introduction to the Primates. Englewood Cliffs, N.J.: Prentice-Hall.

SHOHAMY, E. 1980. Students' Attitudes toward Tests: Affective Considerations in Language Testing. Paper presented at the TESOL '80 Meeting, San Francisco, March. Mimeo.

SINGER, H. 1977. IQ Is and Is not Related to Reading. In S.F. Wanat, (Ed.). Issues in Evaluating Reading. Washington, D.C.: Center for Applied Linguistics.

TROIKE, R. 1981. SCALP: Social and Cultural Aspects of Language Proficiency. Paper presented at the InterAmerica Language Proficiency Assessment Symposium, Airlie House, Virginia, March. To appear in the proceedings.

UPSHUR, J.A. and A. Palmer. 1974. Measures of Accuracy, Communicativity and Social Judgments for two Classes of Foreign Language Speakers. In A. Verdoodt, (Ed.). AILA Proceedings, Copenhagen 1972, Vol. II: Applied Sociolinguistics. Heidelberg: Julius Groos Verlag.

WALKERDINE, V. and C. Sinha. 1981. Developing Linguistic Strategies in Young School Children. In G. Wells. Learning through Interaction. London: Cambridge University Press.

WESCHE, M.B. 1981. Communicative Testing in a Second Language. Canadian Modern Language Review 37.3. (Special issue in honor of H.H. Stern).